Photovoltaic Power Forecasting: Benchmarking Deep Learning Architectures for Scenario-Based Production Scheduling

Assiya Zahid*†, Lamia Hammadi*, Patrice Leclaire†, Roberta Costa Affonso†, Abdessamad El Ballouti*,
*Laboratory of Engineering Sciences for Energy, National School of Applied Sciences, El Jadida, Morocco
†QUARTZ Laboratory, ISAE-SUPMECA, Saint-Ouen, France
Emails: zahid.assiya@ucd.ac.ma, assiya.zahid@isae-supmeca.fr

***Abstract*—Accurate short-term photovoltaic (PV) power forecasting is important for energy-aware production planning in manufacturing environments integrating renewable sources. This paper benchmarks six deep learning architectures—LSTM, BiLSTM, GRU, CNN–LSTM, CNN–BiLSTM, and CNN–GRU—for PV power forecasting at 30-minute resolution under a strictly controlled protocol (identical data split, preprocessing, input window, and evaluation metrics). Experiments on a real annual PV dataset show that CNN–LSTM achieves the best RMSE in the univariate setting (PV power only) with RMSE = 1.153 and $R = 0.972$, while LSTM attains the lowest MAE (MAE = 0.564). In a multivariate extension (PV + weather), CNN–LSTM remains the best hybrid model (RMSE = 1.148, $R = 0.969$), with only marginal RMSE improvement over the univariate CNN–LSTM. Beyond point forecasting, the proposed pipeline converts the best model output into planning-ready uncertainty inputs by generating Monte Carlo trajectories from validation residuals and reducing them into compact multi-day scenarios with probability weights over a weekly horizon. These weighted scenarios are designed to be directly injected as PV availability inputs in uncertainty-aware production scheduling and energy management.**

***Index Terms*—Photovoltaic forecasting, deep learning, CNN–RNN hybrids, uncertainty, clustering, scenario reduction.**

I. Introduction

Solar photovoltaic (PV) systems are increasingly integrated within manufacturing sites as on-site generation assets, driven by declining installation costs and growing pressure on energy-intensive industries to reduce both electricity bills and carbon footprints [1]. In this setting, PV is not merely an additional power source: it becomes an operational lever that reshapes how production and energy are jointly managed.

In manufacturing, energy-aware decisions such as grid purchasing under time-of-use tariffs, battery charging and discharging, load shifting, and production-rate adjustments depend on the expected PV generation [2]. The tighter the coupling between energy management and production scheduling, the more sensitive operations become to the quality of PV forecasts.

However, PV generation is inherently intermittent and uncertain. Output fluctuates with the day–night cycle, seasonal variations, and fast-changing meteorological conditions such as cloud cover and irradiance transients [3]. Unlike dispatchable generation, PV power cannot be controlled on demand and exhibits challenging patterns such as abrupt intra-day ramps and prolonged nighttime zero plateaus, which make forecasting difficult.

This uncertainty propagates directly into production planning: overestimation forces costly peak-hour grid purchases, while underestimation leads to missed savings and fragile schedules [2]. Managing this uncertainty requires accurate and operationally suitable PV forecasts. In industrial practice, planning is rarely based on a single static prediction: decisions are updated under a rolling-horizon paradigm, where day-ahead and intra-day forecasts are repeatedly refreshed and fed into successive optimization steps [4].

In this work, forecasts are produced at 30-minute resolution (48 steps/day) and rolled out iteratively to cover a 7-day operational planning horizon. Moreover, point forecasts alone are often insufficient for robust decision-making: scenario-based scheduling requires an explicit probabilistic representation of PV uncertainty to hedge against adverse outcomes and exploit favorable ones. Transforming forecasts into planning-ready scenario sets is therefore a critical component of PV-integrated industrial planning.

In the targeted production scheduling setting, PV scenarios provide time-indexed realizations of on-site renewable power over the weekly horizon. These trajectories are used as exogenous inputs in uncertainty-aware optimization (e.g., grid purchases under time-of-use tariffs, battery charge/discharge decisions, and feasible production schedules), enabling solutions that hedge against low-generation outcomes while exploiting favorable ones within a rolling-horizon cycle.

To provide such inputs, PV forecasting methods span physical models, statistical approaches, and data-driven models [5]. Statistical approaches such as SARIMA offer interpretability but are limited by linearity and approximate stationarity assumptions, which constrain their ability to capture cloud-driven variability and sharp ramp transitions [6]. This work focuses exclusively on deep learning architectures, which have become increasingly prominent for learning nonlinear temporal dependencies across multiple scales [5]. Despite this progress, two gaps persist in the literature. First, models are often compared under heterogeneous experimental conditions—data splits, preprocessing pipelines, input windows, and metrics—which confounds conclusions about relative performance. Second, evaluation is typically restricted to point-forecast metrics, overlooking how forecast quality translates into scheduling outcomes under uncertainty.

This paper addresses both gaps by presenting: (i) a strictly controlled benchmarking protocol for six deep learning PV forecasters under identical experimental conditions, and (ii) an end-to-end pipeline that converts point forecasts into probabilistic scenario sets for rolling-horizon planning. Forecast residuals drive Monte Carlo trajectory generation, and k-means clustering reduces the ensemble into a compact set of representative scenarios with probability weights, yielding planning-ready inputs for uncertainty-aware scheduling and energy management [7].

The main contributions of this paper are as follows:

- A benchmarking of six deep learning architectures—LSTM, BiLSTM, GRU, CNN–LSTM, CNN–BiLSTM, and CNN–GRU—for PV forecasting at 30-minute resolution (48 steps/day);
- A reproducible evaluation protocol enforcing identical data splits, preprocessing, input window length, and evaluation metrics across all models;
- An analysis of PV-specific time-series behaviors (nighttime zero plateaus and intra-day ramps) and their impact on forecasting errors across model families;
- An end-to-end forecasting-to-scenarios pipeline combining Monte Carlo trajectory generation from forecast residuals with k-means scenario reduction, producing representative scenarios with probability weights for uncertainty-aware optimization in PV-integrated industrial settings.

The remainder of this paper is organized as follows. Section II reviews related work. Section III presents the models, the experimental protocol, and the scenario generation procedure. Results are reported in Section IV, and Section V concludes the paper.

II. Related Work

This section reviews (i) deep learning architectures for PV forecasting and (ii) scenario generation methods used for uncertainty-aware planning in PV-integrated industrial settings.

A. Deep Learning and Hybrid Architectures for PV Forecasting

Recurrent neural networks were among the first deep models applied to PV forecasting. LSTM and GRU architectures typically outperform classical statistical baselines on short-horizon tasks by capturing nonlinear temporal dependencies that ARIMA-family models cannot represent [8]. However, performance often degrades under highly variable irradiance and abrupt ramp events, motivating more expressive hybrid designs.

Bidirectional variants (BiLSTM) extend standard recurrent models by exploiting both past and future context within the input window. For instance, [9] reported comparable performance between BiLSTM and GRU on a grid-connected system in Errachidia (Morocco), and showed that model ranking may depend on evaluation choices (daytime-only vs. full-day including nighttime zero plateaus). Beyond pure autoregressive inputs, feature design is also important: Bamisile et al. [10] highlighted that meteorological and calendar variables, as well as site-specific variability, can strongly affect accuracy and generalization across locations.

Hybrid CNN–recurrent architectures aim to improve robustness by combining local temporal pattern extraction (via 1D-CNN layers) with longer-term sequence modeling (via LSTM/GRU blocks). Attention mechanisms are frequently added to emphasize informative time segments, and several studies report improvements over pure recurrent baselines [11].

Multi-step forecasting remains particularly challenging because uncertainty accumulates over the horizon and ramps may be amplified. Encoder–decoder and attention-enhanced designs are therefore increasingly used for multi-step outputs, sometimes leveraging exogenous weather forecasts when available [12]. Complementary gains are also reported through systematic hyperparameter optimization and hybrid pipelines including smoothing, feature selection, or ensemble strategies [12].

Despite these advances, a recurring limitation is the heterogeneity of experimental conditions across studies—datasets, resolutions, train/test splits, input formulations, and evaluation metrics—which makes cross-study comparisons unreliable and hinders the identification of consistently superior architectures. The present paper addresses this gap through a unified benchmarking protocol applied to all six architectures under strictly identical conditions.

B. Scenario Generation for Uncertainty-Aware Planning

Point forecasting alone is insufficient for robust operational decision-making under renewable uncertainty. In stochastic optimization, uncertain PV output is represented as a discrete set of scenarios, each associated with a probability weight, and propagated through the decision model to obtain solutions that remain effective across plausible realizations [2]. This paradigm is well established in power systems and microgrid management [13], and has been extended to energy-aware manufacturing scheduling where PV uncertainty affects both production and storage decisions.

A common approach to scenario construction is to generate an ensemble of plausible PV trajectories via Monte Carlo simulation driven by empirical forecast residuals [13]. Since large ensembles are computationally expensive for downstream optimization, scenario reduction is typically applied. Clustering methods such as k-means are widely used for scenario reduction on daily profiles or concatenated multi-day trajectories, yielding a compact set of representative scenarios while preserving key temporal shapes and ramp behaviors [7]. The number of retained scenarios balances representativeness and computational tractability, which is particularly critical in rolling-horizon industrial planning where the optimization must be solved repeatedly at each decision epoch.

While these components are individually established, their explicit integration within a controlled PV forecasting benchmark over industrial rolling horizons remains limited. This work addresses the gap by embedding Monte Carlo trajectory generation and k-means-based scenario reduction within a unified evaluation pipeline targeting a 7-day operational horizon representative of industrial scheduling practice.

III. Methodology

This section presents the forecasting benchmark and the pipeline used to generate multi-day PV trajectories over the operational horizon. The workflow includes data preparation with chronological splitting, point-forecast model training, and Monte Carlo trajectory generation with k-means scenario reduction for decision support.

A. Predictive models

1) Recurrent neural networks (RNNs): RNNs are designed to learn temporal dependencies from sequential data (e.g., windowed time series). Classical RNNs may suffer from vanishing gradients when long-term dependencies are required; therefore, this work considers widely used gated and bidirectional variants.

a) Long Short-Term Memory (LSTM): LSTM introduces a memory cell regulated by input, forget, and output gates, enabling the network to retain or discard information over long sequences and mitigating the vanishing-gradient problem. This makes LSTM well-suited for PV power forecasting, where past patterns and evolving conditions influence future outputs.

b) Bidirectional Long Short-Term Memory (BiLSTM): BiLSTM extends LSTM by processing the input sequence in both forward and backward directions. By combining information from past and future contexts within the input window, BiLSTM can improve representation learning and potentially enhance forecasting accuracy when intra-window dynamics are important.

c) Gated Recurrent Unit (GRU): GRU is a simplified gated recurrent architecture that captures long-term temporal dependencies with fewer parameters than LSTM. It relies on update and reset gates, which often leads to faster training and reduced computational cost while maintaining competitive forecasting performance.

2) CNN–RNN hybrid models (Conv1D): To strengthen feature extraction from PV time series, a one-dimensional convolutional front-end (Conv1D) is integrated before the recurrent layers. Conv1D filters slide along the temporal dimension to detect local patterns and short-term fluctuations (e.g., ramps), while a pooling layer (e.g., max pooling) reduces dimensionality and improves efficiency. In this study, this block is combined with recurrent architectures to form CNN–LSTM, CNN–BiLSTM, and CNN–GRU hybrid models, aiming to capture both local temporal structures (CNN) and longer-range dependencies (RNN).

B. Model Architecture and Training Setup

Six deep learning architectures were evaluated in this benchmark: LSTM, BiLSTM, GRU, CNN–LSTM, CNN–BiLSTM, and CNN–GRU. The recurrent baselines use a single recurrent layer with 64 units followed by a Dense(8, ReLU) layer and a linear output. Hybrid models prepend a causal Conv1D layer with 32 filters, kernel size $k = 3$, and ReLU activation. For CNN–BiLSTM and CNN–GRU, a MaxPool(2) layer is additionally applied. The recurrent block contains 64 units and is followed by a Dense(16, ReLU) layer and a linear output.

To ensure a fair comparison, all models were trained under the same experimental protocol using identical data splits, preprocessing pipeline, look-back window length $L = 12$, and evaluation metrics. Two input configurations are considered: a univariate setting with PV power only, and a multivariate setting with $d > 1$ input channels (PV power plus selected weather features), applied to hybrid CNN–RNN models only. This asymmetry is intentional: pure recurrent baselines serve as univariate reference points, while the multivariate extension specifically targets hybrid architectures whose convolutional front-end is better suited to exploit multi-channel inputs.

All models produce a one-step-ahead prediction, corresponding to one 30-minute interval. Multi-day forecasts over the 7-day planning horizon are obtained by recursive rollout: at each step, the model prediction is appended to the input window and the window is advanced by one step. This strategy avoids the need for a dedicated multi-output architecture but may accumulate errors over the horizon.

Training was implemented in Python using TensorFlow/Keras. All models were optimized using Adam with learning rate $\eta = 10^{-4}$ by minimizing the mean squared error (MSE). Training was run for 50 epochs with a mini-batch size of $B = 32$, and performance was monitored using the root mean squared error (RMSE). Early stopping was applied based on validation RMSE, restoring the best weights to mitigate overfitting. The same random seed was used for model initialization and data-window generation. Normalization parameters were fitted on the training set only and then applied to the validation and test sets.

C. Model Evaluation

Model performance is assessed using standard regression metrics that quantify prediction errors and goodness of fit. Let y_i denote the observed value, $\hat{y}_i$ the predicted value, and n the number of samples.

- **Mean Absolute Error (MAE)** measures the average magnitude of the errors:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \, . \tag{1}$$

- **Mean Squared Error (MSE)** penalizes larger errors more strongly by squaring the residuals:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \, . \tag{2}$$

- **Root Mean Squared Error (RMSE)** is the square root of MSE, expressed in the same unit as the target variable:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{3}$$

- **Pearson correlation** (R) measures the linear agreement between observations and predictions:

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}. \tag{4}$$

D. Data Description and Preprocessing

The dataset combines PV power generation and meteorological data from a single ground-mounted 32 MW PV plant located in Minnesota, USA [14]. The forecasting objective targets site-level PV generation, which is the relevant scope for energy-aware production scheduling when a manufacturer integrates its own PV asset.

PV power measurements cover the period from 1 January 2006 to 31 December 2006 and are sampled every 5 minutes. Although this dataset predates more recent deployments, it provides a complete annual cycle with well-documented seasonal variability and is widely used as a reference benchmark in the PV forecasting literature [7], [13]; the benchmarking methodology and pipeline are independent of the specific acquisition year. The series is resampled to a 30-minute resolution by averaging six consecutive 5-minute measurements. Since the target variable is power, each 30-minute value represents the average power over the interval.

Meteorological variables from NREL for the same site and year include global horizontal irradiance (GHI), direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), ambient temperature, relative humidity, and wind speed. These variables are historical measured values, time-aligned with the PV power series; no weather forecasts are used. In operational deployment, numerical weather prediction outputs would be required.

Missing values and outliers are handled before training. When meteorological variables are included, Pearson correlation with PV power is computed for feature screening, as shown in Fig. 1. Pearson coefficients close to $+1$ indicate strong positive linear association with PV power, coefficients close to -1 indicate strong negative association, and values close to zero indicate weak linear association. The retained weather features are GHI, DNI, DHI, and temperature.

As shown in Fig. 2, mid-day PV output is higher and more concentrated in spring and summer, while winter months exhibit lower and more dispersed values, reflecting seasonal irradiance changes relevant to forecasting difficulty.

The dataset is split chronologically into three non-overlapping subsets: 70% for training, 10% for validation, and 20% for testing. This choice strictly preserves temporal causality and prevents information leakage, at the cost of a seasonal asymmetry between subsets—a known limitation of chronological splitting that is standard practice in time-series

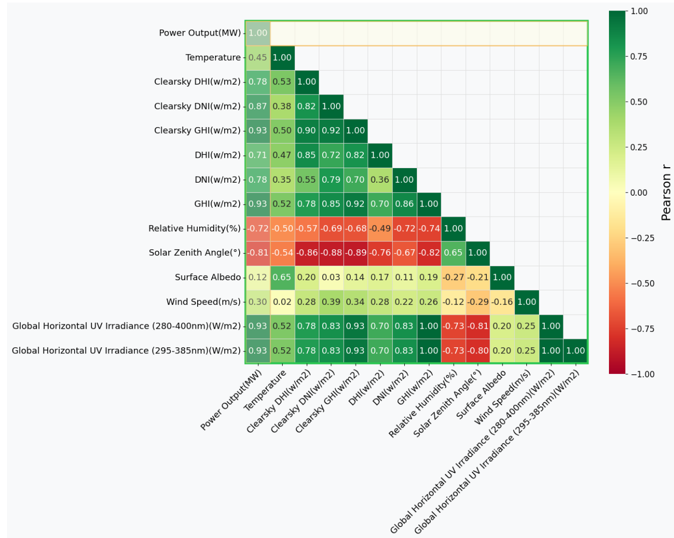

Fig. 1. Correlation heatmap between candidate inputs and PV power output.

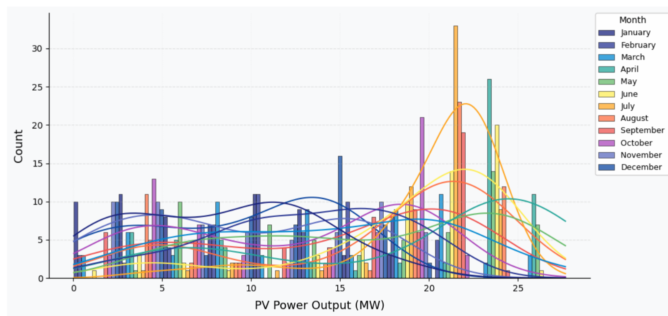

Fig. 2. Monthly distribution of PV power output at 12:00. The count axis represents the number of observations within each PV power interval.

forecasting evaluation. The validation set is used to estimate the residual distribution for Monte Carlo scenario generation, while the test set is used only for final evaluation.

E. Scenario Generation over the Multi-Day Planning Horizon

For scenario-based planning, the best model forecast is converted into probabilistic PV scenarios over a 7-day horizon ($H = 7 \times 48$ steps). This weekly horizon matches short-term production planning cycles and limits recursive error accumulation [2]. The procedure consists of four stages.

Stage 1 — Daily archetype identification Historical PV data are reshaped into daily profiles and clustered using k-means to identify K_{day} representative archetypal patterns capturing recurrent meteorological conditions. The value of K_{day} is selected by running k-means for increasing values (starting from $K_{\mathrm{day}} = 2$) and evaluating, for each partition, the mean absolute percentage error (MAPE) between the historical PV generation duration curve and its cluster-based reconstruction obtained from the representative profiles and their associated weights. To avoid bias from nighttime zero-production plateaus, the duration-curve error is computed over daylight (PV>0) periods only. Each archetype is assigned an empirical occurrence probability estimated from historical cluster frequencies.

Stage 2 — Point forecast over the horizon The selected forecasting model generates a recursive point forecast $\boldsymbol{\mu}$ over the full horizon by iterative rollout. The forecast is assumed reliable over the first $D_{\text{rec}} = 5$ days; for the remaining two days, daily archetypes are used as baseline profiles to mitigate degradation induced by recursive error accumulation.

Stage 3 — Monte Carlo multi-day trajectory generation Forecast residuals are estimated on the validation set and organized into daily error profiles. Let $\tilde{y}_t$ denote the baseline PV profile over the weekly horizon (forecast-driven for $d \le D_{\text{rec}}$ and archetype-driven for $d > D_{\text{rec}}$). Multi-day PV trajectories are then generated by Monte Carlo simulation: for each run $m = 1, \ldots, M$, a residual profile is randomly sampled and used to perturb the baseline profile, yielding

$$y_t^{(m)} = \tilde{y}_t + \varepsilon_t^{(m)}, \qquad m = 1, \ldots, M. \tag{5}$$

Stage 4 — Scenario reduction and probabilities The Monte Carlo ensemble is reduced to a compact set of $K = 10$ representative multi-day scenarios using k-means clustering. For each cluster, the representative scenario s_k is chosen as the closest Monte Carlo trajectory to the cluster centroid (medoid), ensuring physical plausibility. Scenario probabilities are estimated by empirical cluster frequencies:

$$\pi_k = \frac{|\mathcal{C}_k|}{M}, \qquad k = 1, \ldots, K, \tag{6}$$

where $\mathcal{C}_k$ denotes the set of Monte Carlo trajectories assigned to cluster k and M is the total number of simulated trajectories.

IV. Results and Discussion

A. Forecasting Performance

Table I reports the test-set performance of all deep learning models at 30-minute resolution using MAE, MSE, RMSE, Pearson correlation R, and inference time.

a) Univariate benchmark (PV power only): In the univariate setting, the lowest RMSE is achieved by CNN–LSTM (RMSE = 1.152494, R = 0.972347), while the lowest MAE is obtained by LSTM (MAE = 0.563821 MW). BiLSTM reaches a similar correlation (R = 0.972360) but exhibits higher errors (MAE = 0.796065, RMSE = 1.187240). Overall, correlation values remain high and close across models (about 0.969–0.972), indicating that all architectures capture the dominant diurnal pattern; performance differences mainly appear in MAE and RMSE. Among hybrid architectures, CNN–LSTM provides the most consistent improvement in RMSE compared with its recurrent baseline, whereas CNN–BiLSTM and CNN–GRU do not outperform the best univariate baselines. Fig. 3 illustrates the predicted versus actual output of the univariate CNN–LSTM model over the test period.

b) Multivariate extension (PV power + weather): In the multivariate setting (PV power + weather), CNN–LSTM remains the best hybrid model (MAE = 0.652047, RMSE = 1.147945, R = 0.968745). Compared with its univariate counterpart, the multivariate CNN–LSTM yields a

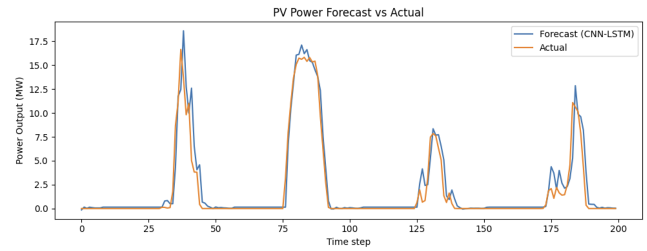

Fig. 3. Univariate CNN–LSTM model: predicted output vs. actual data.

marginal RMSE improvement ($1.152494 \to 1.147945$) but a higher MAE ($0.616422 \to 0.652047$) and a slightly lower correlation ($0.972347 \to 0.968745$). Similar patterns are observed for CNN–BiLSTM and CNN–GRU. This suggests that, under the current configuration (30-min grid, look-back $L = 12$, and the adopted feature set), weather variables provide limited complementary information and may add noise or residual misalignment effects. Univariate inference times (103–108 ms) remain negligible; multivariate hybrids require 341–384 ms, still well within the 30-minute decision step.

c) Error behavior.: Nighttime zero plateaus are generally well handled by all models. Most errors occur during intra-day ramp transitions, especially during morning ramp-up, evening ramp-down, and rapid cloud-driven fluctuations. The better RMSE performance of CNN–LSTM can be explained by its architecture: the Conv1D front-end extracts short local patterns such as ramps, while the LSTM layer captures temporal dependencies across the input window. In contrast, pure recurrent models mainly focus on sequential dependencies and may be less effective at detecting abrupt local variations. Among recurrent-only models, LSTM achieves the lowest MAE, suggesting that its gating mechanism captures the dominant diurnal pattern more effectively under this dataset.

B. Scenario Generation

Scenario generation is based on the univariate CNN–LSTM model, which achieves the lowest RMSE in Table I. The model output is converted into week-ahead PV uncertainty scenarios using $M = 1000$ Monte Carlo trajectories and reduced to $K = 10$ representative scenarios for scheduling purposes. As shown in Fig. 4, the P10–P90 band widens during daylight ramps and peak production periods, where forecast uncertainty is highest, and collapses during nighttime zero-production periods. The reduced scenarios preserve the main intra-day PV patterns and cover three generation regimes: high (peak $P > 15\,\text{MW}$, $\pi \approx 0.17$), intermediate (peak $5 < P < 15\,\text{MW}$, $\pi \approx 0.10$–0.12), and low (peak $P < 5\,\text{MW}$, $\pi \approx 0.06$–0.09), with probabilities estimated from cluster frequencies.

V. Conclusion

This paper presented a controlled benchmark of six deep learning architectures for short-term PV power forecasting at 30-minute resolution, together with an end-to-end pipeline for probabilistic scenario generation over a weekly planning horizon. Under a unified experimental protocol, the best univariate

TABLE I

TEST-SET PERFORMANCE OF ALL BENCHMARKED MODELS AT 30-MINUTE RESOLUTION.

	Univariate (PV power only)						Multivariate (PV power + weather)		
Metric	**LSTM**	**BiLSTM**	**GRU**	**CNN–LSTM**	**CNN–BiLSTM**	**CNN–GRU**	**CNN–LSTM**	**CNN–BiLSTM**	**CNN–GRU**
MAE	**0.563821**	0.796065	0.683470	0.616422	0.689968	0.711002	**0.652047**	0.761804	0.682646
MSE	1.366191	1.409539	1.424275	**1.328242**	1.648118	1.535149	**1.317778**	1.654690	1.553190
RMSE	1.168842	1.187240	1.193430	**1.152494**	1.283791	1.239011	**1.147945**	1.286347	1.246270
R	0.971559	**0.972360**	0.971961	0.972347	0.968833	0.969984	**0.968745**	0.961227	0.962964
Time (ms)	105.055794	107.404416	**103.086891**	108.032852	107.076322	103.518863	383.995323	384.347133	**341.378074**

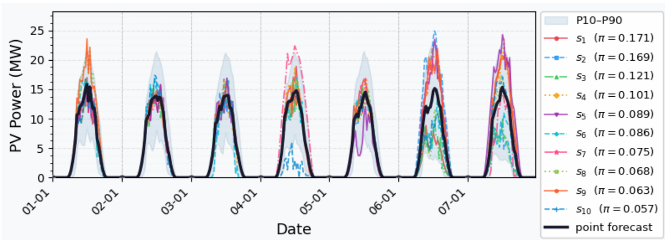

Fig. 4. Weekly PV scenario set derived from the CNN–LSTM forecast, including the P10–P90 band, the point forecast, and the reduced scenarios with probabilities π_k.

RMSE is obtained by CNN–LSTM (RMSE = 1.152 MW, R = 0.972), while the lowest MAE is achieved by LSTM (MAE = 0.564 MW). In the multivariate extension (PV + weather), CNN–LSTM remains the best hybrid model (RMSE = 1.148 MW, R = 0.969), but the gain over the univariate CNN–LSTM is marginal and does not consistently improve all metrics, suggesting limited complementary information from weather inputs under the current configuration. Overall, correlation values remain high and close across models, indicating that all architectures capture the dominant diurnal PV pattern, with performance differences mainly reflected in MAE and RMSE.

Beyond point forecasting, the proposed forecasting-to-scenarios pipeline combines Monte Carlo trajectory generation based on validation residuals with k-means scenario reduction (K = 10), producing compact probabilistic scenario sets with associated probability weights that can be directly used in uncertainty-aware planning and scheduling. This study remains limited by the use of a single-site, one-year dataset, the absence of real industrial deployment, and the lack of explainability analysis. Future work will integrate these scenarios into a two-stage stochastic optimization model for flexible job shop scheduling with PV integration. It will also focus on validating the pipeline in an industrial case study and improving model interpretability through feature-importance and explainability techniques such as SHAP-based analysis.

ACKNOWLEDGMENT

The authors thank the bilateral program PHC TOUBKAL 2024 France-Morocco of CAMPUS FRANCE (Toubkal/24/184 Campus France: 50009RM) and the National Center for Scientific and Technical Research (CNRST) of Morocco in the context of the "PhD Associate Scholarship PASS" program for their support.

REFERENCES

[1] M. Husein, E. Gago, B. Hasan, and M. Pegalajar, "Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies," *Heliyon*, vol. 10, no. 13, 2024.

[2] R. A. Al-Lawati, J. L. Crespo-Vazquez, T. I. Faiz, X. Fang, and M. Noor-E-Alam, "Two-stage stochastic optimization frameworks to aid in decision-making under uncertainty for variable resource generators participating in a sequential energy market," *Applied Energy*, vol. 292, p. 116882, 2021.

[3] S. Karimi and S. Kwon, "Comparative analysis of the impact of energy-aware scheduling, renewable energy generation, and battery energy storage on production scheduling," *International Journal of Energy Research*, vol. 45, no. 13, pp. 18 981–18 998, 2021.

[4] D. Yang and J. Kleissl, *Solar irradiance and photovoltaic power forecasting*. CRC Press, 2024.

[5] N. Khortsriwong, P. Boonraksa, T. Boonraksa, T. Fangsuwannarak, A. Boonsrirat, W. Pinthurat, and B. Marungsri, "Performance of deep learning techniques for forecasting pv power generation: A case study on a 1.5 mwp floating pv power plant," *Energies*, vol. 16, no. 5, p. 2119, 2023.

[6] A. Shadab, S. Ahmad, and S. Said, "Spatial forecasting of solar radiation using arima model," *Remote Sensing Applications: Society and Environment*, vol. 20, p. 100427, 2020.

[7] B. Li, K. Sedzro, X. Fang, B.-M. Hodge, and J. Zhang, "A clustering-based scenario generation framework for power market simulation with wind integration," *Journal of Renewable and Sustainable Energy*, vol. 12, no. 3, 2020.

[8] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep lstm-rnn," *Neural computing and applications*, vol. 31, no. 7, pp. 2727–2740, 2019.

[9] A. Ait Mansour, A. Tilioua, and M. Touzani, "Bi-lstm, gru and 1d-cnn models for short-term photovoltaic panel efficiency forecasting case amorphous silicon grid-connected pv system," *Results in Engineering*, vol. 21, p. 101886, 2024.

[10] O. Bamisile, A. Oluwasanmi, S. Obiora, E. Osei-Mensah, G. Asoronye, and Q. Huang, "Application of deep learning for solar irradiance and solar photovoltaic multi-parameter forecast," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 46, no. 1, pp. 13 237–13 257, 2024.

[11] L. Zhuo and X. Hejin, "Photovoltaic output prediction based on cnn-gru-attention algorithm," in *2023 China Automation Congress (CAC)*. IEEE, 2023, pp. 4309–4313.

[12] Y. Dai, W. Yu, and M. Leng, "A hybrid ensemble optimized bigru method for short-term photovoltaic generation forecasting," *Energy*, vol. 299, p. 131458, 2024.

[13] M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an lstm neural network and synthetic weather forecast," *Ieee Access*, vol. 8, pp. 172 524–172 533, 2020.

[14] National Renewable Energy Laboratory, "Solar Power Data for Integration Studies," https://www.nrel.gov/grid/solar-power-data.html, 2006, 32 MW PV plant, Minnesota, USA. Accessed: 2024.